Journal of Medical Virology 57:68–74 (1999)

Phylogenetic Analysis of Hepatitis E Virus Isolates From Egypt†

Sergei A. Tsarev,[1]* Leonard N. Binn,[1] Peter J. Gomatos,[2] Ray R. Arthur,[3] Mohamed K. Monier,[3] Helene van Cuyck-Gandre,[4] Charles F. Longer,[1] and Bruce L. Innis[1]

[1]*Department of Virus Diseases, Walter Reed Army Institute of Research, Washington, DC*
[2]*U.S. Naval Medical Research Unit No. 3, Cairo, Egypt*
[3]*Abbassia Fever Hospital, Cairo, Egypt*
[4]*Centre de Recherches du Service de Sante des Armees, Unite de Virologie, La Tronche, France*

Hepatitis E virus (HEV) genome was detected by reverse transcriptase-polymerase chain reaction (RT-PCR) in fecal samples of two sporadic cases of hepatitis E in Cairo Egypt. Sequence of the complete putative structural region [open reading frame (ORF)-2] and complete region of unknown function (ORF-3) was determined for the two HEV isolates. Phylogenetic analysis of the nucleotide sequences was performed using neighbor joining or maximum parsimony methods of tree reconstruction. Direct correspondence between the HEV evolutionary trees and geographic origin of the HEV isolates was observed. Three genotypes of HEV were identified: genotype I (Asia-Africa), genotype II (US), and genotype III (Mexico). Genotype I was further divided into two subgenotypes (Asia and Africa). In the Asian subgenotype, three smaller genetic clusters were observed (China-like sequences, Burma-like sequences, and sequence from a fulminant case of HEV). The segregation of all these genetic clusters was supported by the high level of bootstrap probabilities. Four regions of the HEV genome were used for phylogenetic analysis. In all four regions, Egyptian HEV isolates were grouped in a separate African clade. ***J. Med. Virol. 57:68–74, 1999.***

Published 1999 Wiley-Liss, Inc.

KEY WORDS: HEV genome; polymerase chain reaction; neighbor joining method; maximum parsimony; bootstrap

INTRODUCTION

Hepatitis E virus (HEV) is a single-stranded positive sense RNA virus. The virus is transmitted by the fecal-oral route and causes hepatitis E, an acute self-limited liver disease that is a major health problem in developing countries. HEV is currently classified in the family Caliciviridae. However, significant differences of its genome organization from that of other caliciviruses [Koonin et al., 1992] might require HEV reclassification. HEV can cause severe disease especially in pregnant women; therefore, a vaccine for hepatitis E is needed. Recombinant DNA technology has been used successfully to create diagnostic tests and vaccine candidates for hepatitis E [Purdy et al., 1993; Tsarev et al., 1994]. Most hepatitis E cases are reported from Asia; therefore, candidate vaccines were based on a structural HEV protein derived from Asian strains. Consequently, analysis of HEV sequences from different geographic origins is important.

Several HEV genome sequences were determined recently, enabling comparative phylogenetic analysis. Phylogenetic clustering was found to be geographically related. HEV sequences from Asia are distinguished from those of Africa despite a close relationship [Chatterjee et al., 1997; van Cuyck-Gandre et al., 1997]. The single isolate from Mexico was the most genetically distant isolate [Huang et al., 1992]. Two HEV isolates from the United States (US) were described recently which differed from the Asia-African group to the same extent as from the Mexican HEV strain [Meng et al., 1997; Schlauder et al., 1998]. There is also evidence that significant heterogeneity of HEV might be present within one country. In the limited genome region available for evaluation, two of seven HEV isolates from China had unexpectedly large differences from the con-

The views presented are those of the authors and do not purport to reflect those of the U.S. Department of the Army, U.S. Department of the Navy, U.S. Department of Defense, or Ministry of Defense, France.

The current address of Ray R. Arthur is Division of Emerging and Other Communicable Diseases, World Health Organization, Geneva, Switzerland.

*Correspondence to: Dr. Sergei A. Tsarev, Department of Virus Diseases, Walter Reed Army Institute of Research, Building 40, Room 2042, Washington, DC 20307. E-mail: TSAREVS@WRSMTP-CCMAIL.ARMY.MIL

Accepted 17 July 1998

PUBLISHED 1999 WILEY-LISS, INC.

sensus Asia-African sequences [Huang et al., 1995]. Lacking techniques for meaningful antigenic analysis, we must evaluate the extent of the HEV heterogeneity by analyzing sequences from diverse geographical areas. Here we report complete open reading frame (ORF)-3,2 regions of two HEV isolates from sporadic cases in Egypt, their phylogenetic relationships to other HEV strains, and a statistical analysis for the observed tree topology.

MATERIALS AND METHODS

Fecal Samples

Fecal samples of two sporadic cases of hepatitis E (93-Egypt and 94-Egypt) which occurred in Cairo in 1993/1994 were collected on the 7th and 9th days after the onset of jaundice. Twenty percent phosphate-buffered saline (PBS) fecal suspensions were kept at −70°C. The specimens tested negative in cell culture for other enteric viruses.

HEV 93-Egypt was recovered from a 23-year-old male army recruit living in a rural area outside of Cairo. Seven days after onset of jaundice, the patient complained of malaise, nausea, vomiting, anorexia, dark urine, and abdominal pain. On examination, he was febrile and had an enlarged liver. Previously he had schistosomiasis. Laboratory study revealed: AST 1,100 U/l, ALT 1,124 U/l, total bilirubin 15.1 mg/dl, and IgM and IgG to HEV.

HEV 94-Egypt was recovered from a 26-year-old male army recruit living in an urban area in Cairo. Nine days after the onset of jaundice, the patient complained of malaise, nausea, anorexia, dark urine, arthralgia, and abdominal pain. On examination, he was febrile and had an enlarged liver. Laboratory tests revealed: AST 392 U/l, ALT 500 U/l, total bilirubin 5.7 mg/dl, and IgM and IgG to HEV.

RNA Extraction and Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

RNA was extracted with TRIZol reagent (GIBCO/BRL, Gaithersburg, MD) from 100 μl of 20% fecal suspension in accordance with the manufacturer's protocol.

RT and nested PCR steps were performed as described earlier [Tsarev et al., 1994].

PCR fragments were purified from 100 μl of PCR mixture with QIAquick PCR Purification Kit (Qiagen, Chatsworth, CA) and directly sequenced by automated DNA sequencing on ABI Model 373.

More than 95% of the sequences were determined from both strands.

Sequence Analysis

Sequences were assembled using Sequencher 3.0 (Gene Codes Corporation, Ann Arbor, MI) and analyzed with the Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, on an Alpha 8400 computer (Unix, Frederick Biomedical Supercomputing Center). The numbers of synonymous (silent) and non-synonymous substitutions were calculated with the program SILENT written by Smythers [1996] for the VAX 662 (VMS). Multiple sequence alignments were performed with GCG-PileUp program.

A consensus evolutionary tree was produced with GCG-PAUPsearch program using neighbor joining or maximum parsimony methods. Confidence for the grouping in the trees was assessed by bootstrap method (1,000 replicates). Bootstrap values of ⩾75% were regarded as providing evidence for the phylogenetic grouping [Zharkikh and Li, 1992]. Non-consensus evolutionary trees were produced by GCG-GROWTREE. The graphic output of the phylogenetic trees was created with the Macintosh version of the TREEVIEW program [Page, 1996].

Nucleotide Sequence Accession Numbers

The sequences reported in this article have been deposited in the GenBank data base. Accession numbers of the sequences used for analysis in this study are shown in Table I.

RESULTS

Sequencing of the 93-Egypt and 94-Egypt HEV Isolates

Two regions of the HEV genome have practical importance: ORF-3 of unknown function and the capsid region ORF-2. Fragments of ORF-3 and ORF-2 expressed in *Escherichia coli* are currently used in commercial serological tests. Proteins expressed from the ORF-2 region were also used for two vaccine candidates against hepatitis E [Purdy et al., 1993; Tsarev et al., 1994]. Therefore, in this study of 93-Egypt and 94-Egypt (two sporadic HEV isolates from Africa), we determined the complete sequences of the ORF-3 and ORF-2.

To enable comparison with other African isolates, an additional sequence of about 100 nucleotides (nt) was determined from the non-structural region ORF-1.

Complete ORF-3 and ORF-2 regions of these isolates were compared to other known HEV sequences (Table II). Because in two other genomic regions, ORF-1,2,3 overlap and the 3′-end of ORF-2, more African sequences are available, they were also included in the comparison.

The majority of nucleotide changes observed in the HEV genome are silent changes. Because most of the ORF-3 region overlaps with the ORF-2 region, few silent changes are possible in ORF-3. As expected, this region has the lowest mean proportion of silent nucleotide changes compared to the overall number of nucleotide changes. Despite the low mean proportion of silent mutations, the range and standard deviation for this parameter are large. Restriction of the number of silent nucleotide changes makes ORF-3 one of the most conserved regions of the HEV genome at the nucleotide level. Homology between the two Egyptian isolates and other HEV isolates was the highest in this region. The nucleotide homology was decreased in the following sequence: complete ORF-3, ORF-1,2,3 overlap, complete ORF-2, and ORF-2,3′-end.

TABLE I. HEV Sequences Used for Analysis

Name	Abbreviation	GenBank accession no.	Reference
82[a]-Burma	82-Bur	M73218	Tam et al. [1991]
89-Myanmar (Burma)	89-Mya	D10330	Aye et al. [1993]
90-India (Hyderabad)	90-Ind	U22532	Panda et al. [1995]
93-India (Madras)	93-Ind	X99441	Unpublished data
87-Nepal	87-Nep	AF020495	Gouvea et al. [1997]
94-Nepal	94-Nep	AF020492	Gouvea et al. [1997]
95-Nepal	95-Nep	AF020497	Gouvea et al. [1997]
87-China-A	87-Chi-A	D11092	Aye et al. [1992]
87-China-B	87-Chi-B	M94177	Bi et al. [1993]
87-China-C	87-Chi-C	L25595	Yin et al. [1994]
87-China-D	87-Chi-D	D11093	Unpublished data
87-Pakistan	87-Pak	M80581	Tsarev et al. [1992]
88-Kirgizia	88-Kirg	AF010417	Chatterjee et al. [1997]
88-Uzbekistan	88-Uzb	AF010422	Chatterjee et al. [1997]
92-Fulminant (India?)	92-Fulm	X98292	Donati et al. [1997]
80-Algeria	80-Alg	U40046, AF001275	van Cuyck-Gandre et al. [1997]
		AF051350	This paper
83-Tunisia	83-Tun	AF010421	Chatterjee et al. [1997]
84-Chad	84-Chad	U62121, AF001276	van Cuyck-Gandre et al. [1997]
		AF051349	This paper
93-Egypt	93-Egy	AF051351	This paper
94-Egypt	94-Egy	AF051352	This paper
94-Morocco	94-Mor	AF010418	Chatterjee et al. [1997]
87-Mexico	87-Mex	M74506	Huang et al. [1992]
95-USA (human)	95-USA-H	AF035437	Schlauder et al. [1998]
96-USA (swine)	96-USA-S	AF011921	Meng et al. [1997]

[a]Year of isolation. When an epidemic lasted more than 1 calendar year and the year of the isolate was not clear from a publication, the isolate was assigned the middle or the last year of that epidemic.

These four regions of the HEV genome were used for the phylogenetic analysis. However, only in the two ORF-2 regions was an acceptable level of statistical support for the tree topology achieved (Figs. 1, 2). No statistical support was obtained for the topology of evolutionary trees produced for the overlap ORF-1,2,3 or complete ORF-3 regions. Nevertheless, for comparison with previously published data, the non-consensus tree of the overlap region also is presented here (Fig. 3). The high level of amino acid conservation in the four regions (Table II) makes amino acid sequences unsuitable for phylogenetic analysis.

Phylogenetic Analysis of the Complete HEV ORF-2 Region

We previously have demonstrated that HEV isolates could be separated into four genetic groups by comparison of short sequences from different parts of HEV genome [Chatterjee et al., 1997; van Cuyck-Gandre et al., 1997]. However, the phylogenetic trees obtained were not subjected to statistical analysis. In this study, we apply the bootstrap statistical test to evaluate the validity of the phylogenetic tree topology. We utilized the two most commonly used methods of tree construction: neighbor joining (NJ) method and maximum parsimony (MP) method. Understanding of the statistical meaning of bootstrap values has undergone several revisions since first introduced into phylogeny by Felsenstein [1985]. A cutoff value of approximately 75% is generally accepted now to establish confidence in phylogenetic grouping [Zharkikh and Li, 1992].

The bootstrap consensus tree produced for the complete ORF-2 region of all currently available HEV isolates is shown in Figure 1. Lemon and Robertson [1993] proposed to classify phylogenetic groups with genetic distances greater than 15% as genotypes. Three genotypes are observed according to this criterion in the complete ORF-2 region: genotype I, which includes Afro-Asian sequences, genotype II, which includes US sequences, and genotype III, the only representative of which is Mexican sequence. The same authors suggested that genetic groups within genotypes which differ from others by more than 7.5% be called subgenotypes. Genotype I consists of two subgenotypes: subgenotype I-1, which consists of sequences from Asia, and subgenotype I-2, which includes sequences from Africa. However, distinct segregation of China-like sequences from Burmese-like sequences and a fulminant sequence were observed in the Asian subgenotype with genetic distances greater than 5.5%. These smaller genetic clusters were called I-1a, I-1b, and I-1c, respectively. All mentioned genetic groups with different level of divergence from each other were formed by both NJ and MP methods with 100% level of confidence, except fulminant and Mexican strains, which are both single members in their groups and therefore no level of confidence could be assigned for their segregation.

In the ORF-2 region, the fulminant sequence fails to group with either of the two Asian subgroups as determined by NJ or MP bootstrap methods. Similar results were observed when we used complete HEV sequences

TABLE II. Nucleotide and Amino Acid Comparison of 93-Egypt and 94-Egypt Isolates to Other HEV Strains

	% of nucleotide (amino acid) identity [% of synonymous nucleotide differences]																			
	ORF-1,2,3 overlap[a] (232 nt)		Complete ORF-3 (372 nt)						Complete ORF-2 (1983)						ORF-2 3′-end (424 nt)					
Isolate	93-Egy	94-Egy	93-Egy			94-Egy			93-Egy			94-Egy			93-Egy			94-Egy		
82-Bur	95	96	99	(100)	[100]	98	(100)	[100]	91	(99)	[96]	91	(99)	[95]	88	(99)	[96]	89	(99)	[96]
89-Mya	94	96	98	(98)	[71]	98	(98)	[64]	91	(99)	[94]	91	(98)	[93]	87	(98)	[93]	88	(98)	[92]
90-Ind	94	93	97	(95)	[50]	95	(95)	[67]	90	(98)	[92]	90	(97)	[90]	87	(97)	[89]	87	(97)	[89]
93-Ind	94	95	98	(100)	[100]	98	(100)	[100]	91	(98)	[94]	91	(98)	[93]	88	(99)	[96]	89	(99)	[96]
87-Nep	—[b]	—		—			—			—			—		88	(99)	[96]	88	(99)	[96]
94-Nep	—	—		—			—			—			—		88	(99)	[96]	89	(99)	[96]
95-Nep	—	—		—			—			—			—		88	(99)	[96]	89	(99)	[96]
87-Chi-A	95	95	99	(100)	[100]	98	(100)	[100]	92	(99)	[96]	92	(99)	[95]	89	(99)	[94]	89	(99)	[94]
87-Chi-B	94	94	98	(98)	[71]	97	(98)	[82]	91	(99)	[94]	91	(98)	[93]	89	(99)	[91]	89	(99)	[91]
87-Chi-C	94	94	98	(98)	[71]	97	(98)	[82]	91	(99)	[94]	91	(98)	[92]	88	(99)	[92]	88	(99)	[92]
87-Chi-D	94	94	98	(98)	[64]	97	(98)	[75]	91	(98)	[92]	91	(98)	[91]	90	(99)	[90]	89	(99)	[90]
87-Pak	94	94	99	(100)	100]	98	(100)	100]	91	(99)	[95]	92	(99)	[93]	88	(99)	[92]	88	(99)	[92]
88-Kirg	95	96		—			—			—			—			—			—	
88-Uzb	95	96		—			—			—			—			—			—	
92-Fulm	94	96	97	(98)	[83]	97	(98)	[83]	91	(98)	[93]	90	(98)	[92]	90	(99)	[93]	90	(99)	[93]
80-Alg	96	96	99	(99)	[67]	98	(99)	[86]		—			—		91	(99)	[92]	90	(99)	[93]
83-Tun	97	97		—			—			—			—			—			—	
84-Chad	97	97	99	(100)	[100]	98	(100)	[100]		—			—		94	(99)	[92]	92	(99)	[94]
93-Egy	100	97		N/A[c]		98	(100)	[100]		N/A		96	(99)	[92]		N/A		96	(99)	[88]
94-Egy	97	100	98	(100)	[100]		N/A		96	(99)	[92]		N/A		96	(99)	[88]		N/A	
94-Mor	95	96		—			—			—			—			—			—	
87-Mex	86	85	91	(87)	[51]	90	(87)	[55]	81	(95)	[84]	81	(94)	[83]	77	(93)	[84]	77	(93)	[84]
95-USA-H	86	85	87	(85)	[51]	86	(85)	[54]	80	(92)	[81]	79	(92)	[81]	80	(92)	[79]	78	(92)	[80]
96-USA-S	—	—	85	(83)	[55]	85	(83)	[56]	79	(93)	[81]	80	(93)	[82]	80	(94)	[82]	80	(94)	[81]
Average	90	90	96	(96)	[76]	96	(96)	[82]	89	(98)	[91]	89	(97)	[90]	88	(98)	[91]	88	(98)	[91]
± SD[d]	± 3	± 3	± 5	± 6	± 20	± 4	± 6	± 17	± 5	± 2	± 5	± 5	± 2	± 5	± 4	± 2	± 5	± 4	± 2	± 5

[a]Only nucleotide identity is shown for the overlap region.
[b]Data not available.
[c]Not applicable.
[d]Standard deviation.

for the MP bootstrap analysis. However, when NJ bootstrap method was applied, the fulminant case was grouped into Chinese subgroup I-1a as the most distant member (data not shown).

It is interesting to note that in the ORF-2 region, the fulminant HEV sequence has the highest homology (94.1%) with the 82-Burma strain and the second highest homology (94.0%) with the 87-China-B strain. In the ORF-1 region, the highest homology for the fulminant strain was observed with 87-China-C isolate (98.7%), while homology with the 82-Burma isolate was significantly lower (93.3%).

Both US sequences derived from human and swine belong to the same genotype. Genetic distance of 8.0% between the two sequences might indicate that they come from different subgenotypes. However, more sequences from US are required for a more definite conclusion.

Egyptian strains are the only two African strains for which the complete ORF-2 sequences were determined. The two Egyptian sporadic cases were separated by about half a year and both cases came from residents of the Cairo area. Therefore, it was surprising to see a moderately high level of genetic divergence between the two strains (96% identity).

Phylogenetic Analysis of the 3′-End HEV ORF-2 Region

The 3′-end region was chosen for analysis because in addition to the two HEV sequences of the Egyptian strains, sequences from African strains from Algeria and Chad and sequences from Nepal are also available in this region [Gouvea et al., 1997; van Cuyck-Gandre et al., 1997]. We were able to perform meaningful statistical analysis for the 3′-end ORF-2 region. The topology of the consensus tree obtained in this region (Fig. 2) was similar to that obtained for the complete ORF-2 (Fig. 1). All African isolates were grouped into the same cluster. The two most closely related sequences in this group were Egyptian strains (96% nucleotide identity). The Egyptian strains were more closely related to the Chad isolate (92% identity) than to the Algerian isolate (90% identity). This segregation of the African strains was statistically significant by NJ but not by MP analysis. The level of divergence between African sequences might indicate that they could be separated into several subgenotypes. However, the current sampling is not large enough to observe distinct clusters.

A tendency for the fulminant isolate to segregate

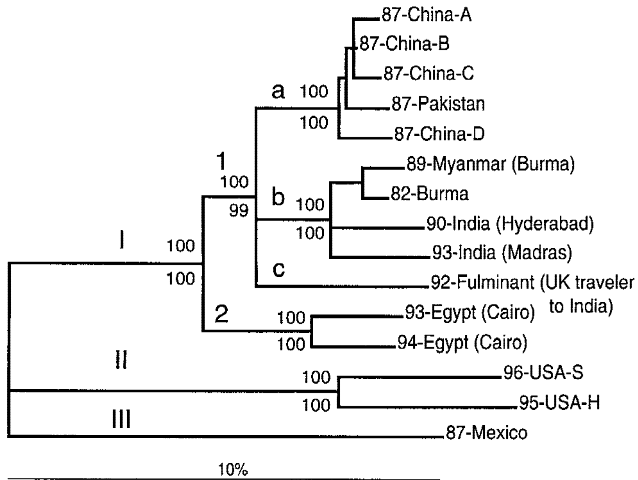

Fig. 1. Bootstrap 50% majority rule consensus phylogenetic tree of HEV isolates based on the complete ORF-2 (1983 nt). The tree was generated from the PileUp alignment of 15 HEV sequences including 2 sequences reported in this study. Genbank accession numbers for all sequences are reported in Materials and Methods. The internal node numbers represent the bootstrap values (expressed as percentage of all trees) obtained from 1,000 replicates. Numbers above the nodes are obtained with neighbor joining bootstrap analysis. Numbers below the nodes are obtained with parsimony bootstrap analysis. Only significant bootstrap probabilities of ≥75% cutoff value are shown. Branch lengths are proportional to the evolutionary distance between sequences. A scale in percent of genetic distances is shown. Roman numbers are used to identify genotypes (>15% nucleotide non-identity), arabic numbers to identify subgenotypes (>7.5% non-identity), and letters to denote less different but distinct phylogenetic clusters within subgenotypes.

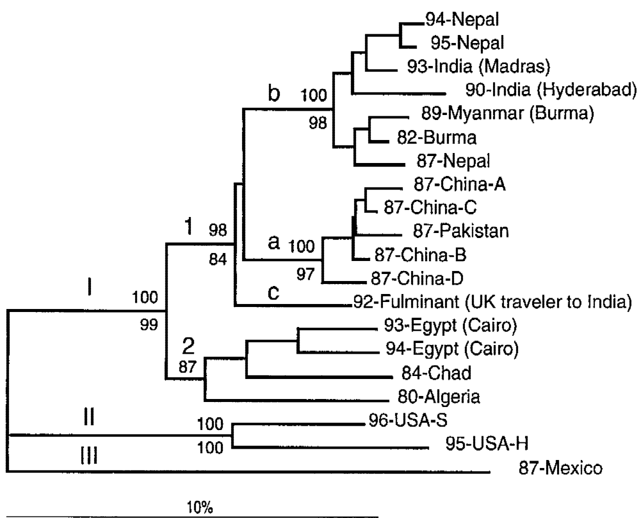

Fig. 2. Bootstrap 50% majority rule consensus phylogenetic tree of HEV isolates based on the ORF-2 3′-end (424 nt). A scale in percent of genetic differences is shown.

from the Asian group was observed, however, without significant bootstrap probability.

Phylogenetic Analysis of the ORF-1,2,3 Overlap Region

Several additional HEV sequences that were unavailable in the ORF-2 region were available for the ORF-1,2,3 overlap region. Most of the clusters observed in the ORF-2 analysis failed to emerge in the bootstrap 50% majority rule trees. Therefore, to be able to compare HEV tree topology with the previously published trees, we used the same distance method of the non-consensus tree reconstruction as was described [Chatterjee et al., 1997; van Cuyck-Gandre et al., 1997] (Fig. 3). Although the level of genetic differences in the ORF-1,2,3 region was lower than in the ORF-2 region, very similar tree topology was observed.

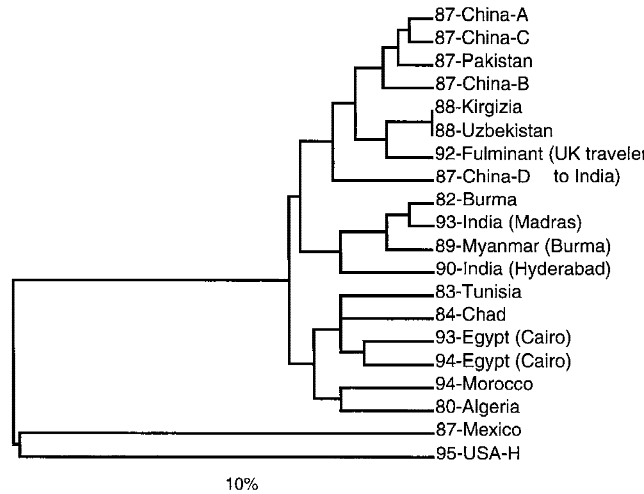

Fig. 3. Phylogenetic tree of HEV isolates in the ORF-1,2,3 overlap region (232 nt). The tree was reconstructed with unweighted pair group method using arithmetic averages. Branch lengths are proportional to the evolutionary distances between sequences. A scale in percent of genetic differences is shown.

All five known African sequences were clustered together. As in the ORF-2 region, the two Egyptian strains were closely related to the Chad strain, and to a new strain from Tunisia, for which sequence in the ORF-2 was unavailable. The more divergent Algerian strain also had more differences in the overlap region and was grouped together with a strain from Morocco.

The fulminant isolate was most closely related to the isolates from Central Asia. However, this grouping was not observed in other genomic regions for which HEV sequences from the Central Asia are available (data not shown).

DISCUSSION

In this study, we confirmed clustering in the evolutionary tree of HEV isolates corresponding to the place of isolation. Three HEV genotypes corresponding to distinct geographic regions were identified: genotype I (Asia-Africa), genotype II (US), and genotype III (Mexico). Most of the known HEV sequences belong to genotype I. This genotype was divided into smaller phylogenetic groups. Two subgenotypes I-1 (Asia) and I-2 (Africa) were segregated with a high level of confidence. Furthermore, the Asian group was divided into three smaller genetic clusters: I-1a (China-like), I-1b

(Burma-like), and I-1c (fulminant case). The geographic origin of the fulminant case (I-1c) is unclear. The genetic groupings of the HEV strains reflected the geographic relationships of the places from which they were isolated. Segregation of the groups was confirmed by bootstrap analysis.

There was a good correlation between the two methods of phylogenetic reconstruction based on distances (NJ) or on MP algorithm. Generally, the parsimony approach was more conservative. There was also reasonable agreement in tree topology derived from the different regions of the HEV genome. The one exception was the HEV from a fulminant case. This strain was isolated from a UK resident who traveled to India [Donati et al., 1997]. It is unclear where in this patient's trip HEV infection occurred. Significant differences in placement of this isolate by the two phylogenetic methods were observed. This strain also demonstrated different relationships to other HEV sequences in different parts of the genome. This might be related to the unique geographic origin of this strain. The HEV from the fulminant case is definitely most closely related to the Asian subgenotype, yet it might represent another subgenotype of the Asia-African genotype. Additional strains from Asia must be evaluated to clarify the tissue. Another explanation of a special place for the fulminant case HEV in the phylogenetic trees might be that there are two types of HEV strains circulating in Asia. One type is more pathogenic and causes fulminant hepatitis more frequently. To test this hypothesis, more strains from fulminant cases of HEV from Asia and other areas must be genetically characterized.

Because the bootstrap probability depends on the length of the region that is analyzed, the most robust tree was obtained for the complete ORF-2 among the four genomic regions analyzed. The segregation of the Asian group into two subgroups was shown at a high level of confidence. The segregation of the African cluster from the Asian one was also at a high level of probability. Bootstrap values obtained for the 3′-end of ORF-2 were also very significant. Egyptian strains were placed into the African clade with other African strains.

In the overlap region, no support for the tree topology was provided by the bootstrap analysis. However, even in the trees with weak topology, African strains always were clustered together, separate from HEV strains of other geographic origin, although with less genetic distance. It appears that the distance method was more sensitive than the parsimony method of tree reconstruction in this case, because this topology was observed only with NJ but not with MP method.

The observed correlation between the evolutionary tree and geography is consistent with a hypothesis that HEV reservoir is stably maintained in geographically restricted regions. Animals might comprise such a reservoir. This hypothesis was directly supported by Clayson et al. [1995] when HEV sequence was detected by PCR in the Nepalese pigs. It was later shown that the human HEV sequences from Nepal were very close to those isolated from pigs (unpublished data). The same situation was observed in the US: HEV sequence isolated from a human acute case of hepatitis E [Schlauder et al., 1998] was very closely related to a sequence isolated from a US pig [Meng et al., 1997].

The geographical segregation of HEV sequences was observed only on the nucleotide level. There is a high level of conservation at the amino acid level. This finding is consistent with the hypothesis that only one serotype of HEV exists. The two genetically most divergent strains from a prototype anti-HEV vaccine based on 87-Pakistan sequence [Tsarev et al., 1994] are Mexican [Huang et al., 1992] and US [Meng et al., 1997] isolates. However, it was demonstrated that this vaccine candidate based on an Asian HEV isolate equally protects animals challenged with the Mexican HEV isolate or homologous virus [Tsarev et al., 1997].

Overall, in spite of the wide spectrum of genetic heterogeneity in HEV, all data indicate that only a monovalent vaccine (derived from a single clade) is required. Moreover, no phenotypic characteristics, such as virulence in humans or animals, have been linked to genetic variations.

ACKNOWLEDGMENTS

We acknowledge the National Cancer Institute for allocation of computing time and staff support at the Frederick Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center.

This study was supported in part by SmithKline Beecham Biologicals through a Cooperative Research and Development Agreement with the U.S. Army Medical Research and Material Command.

REFERENCES

Aye TT, Uchida T, Ma XZ, Iida F, Shikata T, Zhuang H, Win KM. 1992. Complete nucleotide sequence of a hepatitis E virus isolated from the Xinjiang epidemic (1986–1988) of China. Nucleic Acids Res 20:3512.

Aye TT, Uchida T, Ma X, Iida F, Shikata T, Ichikawa M, Rikihisa T, Win KM. 1993. Sequence and gene structure of the hepatitis E virus isolated from Myanmar. Virus Genes 7:95–109.

Bi SL, Purdy MA, McCaustland KA, Margolis HS, Bradley DW. 1993. The sequence of hepatitis E virus isolated directly from a single source during an outbreak in China. Virus Research 28:233–247.

Chatterjee R, Tsarev S, Pilot J, Coursaget P, Emerson S, Purcell R. 1997. African strains of hepatitis E virus that are distinct from Asian strains. J Med Virol 53:139–141.

Clayson ET, Innis BL, Myint KS, Narupiti S, Vaughn DW, Giri S, Ranabhat P, Shrestha MP. 1995. Detection of hepatitis E virus infections among domestic swine in the Kathmandu Valley of Nepal. Am J Trop Med Hyg 53:228–232.

Donati MC, Fagan EA, Harrison TJ. 1997. Sequence analysis of full length HEV clones derived directly from human liver in fulminant hepatitis E. In: Rizetto M, Purcell RH, Gerin JL, Verme G, editors. Viral Hepatitis and Liver Disease. Torino: Edizoni Minerva Medica, p 313–316.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Gouvea V, Snellings N, Cohen SJ, Warren RL, Myint KSA, Shrestha MP, Vaughn DW, Hoke CH Jr, Innis BL. 1997. Hepatitis E virus in Nepal: Similarities with the Burmese and Indian variants. Virus Res 52:87–96.

Huang CC, Nguyen D, Fernandez J, Yun KY, Fry KE, Bradley DW, Tam AW, Reyes GR. 1992. Molecular cloning and sequencing of

the Mexico isolate of hepatitis E virus (HEV). Virology 191:550–558.

Huang R, Nakazono N, Ishii K, Kawamata O, Kawaguchi R, Tsukada Y. 1995. Existing variations on the gene structure of hepatitis E virus strains from some regions of China. J Med Virol 47:303–308.

Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: Delineation of an additional group of positive-strand RNA plant and animal viruses. Proc Nat Acad Sci USA 89:8259–8263.

Lemon SM, Robertson BH. 1993. Current perspectives in the virology and molecular biology of hepatitis A virus. Semin Virol 4:285–295.

Meng XJ, Purcell RH, Halbur PG, Lehman JR, Webb DM, Tsareva TS, Haynes JS, Thacker BJ, Emerson SU. 1997. A novel virus in swine is closely related to the human hepatitis E virus. Proc Nat Acad Sci USA 94:9860–9865.

Page RDM. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358.

Panda SK, Nanda SK, Zafrullah M, Ansari IH, Ozdener MH, Jameel S. 1995. An Indian strain of hepatitis E virus (HEV): Cloning, sequence, and expression of structural region and antibody responses in sera from individuals from an area of high-level HEV endemicity. J Clin Microbiol 33:2653–2659.

Purdy MA, McCaustland KA, Krawczynski K, Spelbring J, Reyes GR, Bradley DW. 1993. Preliminary evidence that a trpE-HEV fusion protein protects cynomolgus macaques against challenge with wild-type hepatitis E virus (HEV). J Med Virol 4:90–94.

Schlauder GG, Dawson GJ, Erker JC, Kwo PY, Knigge MF, Smally DL, Rosenblatt JE, Desai SM, Mushahwar IK. 1998. The sequence and phylogenetic analysis of a novel hepatitis E virus isolated from a patient with acute hepatitis reported in the United States. J Gen Virol 79:447–456.

Smythers GW. 1996. SILENT. Frederick Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center, NCI/SAIC.

Tam AW, Smith MM, Guerra ME, Huang CC, Bradley DW, Fry KE, Reyes GR. 1991. Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome. Virology 185:120–131.

Tsarev SA, Emerson SU, Reyes GR, Tsareva TS, Legters LJ, Malik IA, Iqbal M, Purcell RH. 1992. Characterization of a prototype strain of hepatitis E virus. Proceedings of the National Academy of Sciences of the United States of America 89:559–563.

Tsarev SA, Tsareva TS, Emerson SU, Govindarajan S, Shapiro M, Gerin JL, Purcell RH. 1994. Successful passive and active immunization of cynomolgus monkeys against hepatitis E. Proc Nat Acad Sci USA 91:10198–10202.

Tsarev SA, Tsareva TS, Emerson SU, Govindarajan S, Shapiro M, Gerin JL, Purcell RH. 1997. Recombinant vaccine against hepatitis E—Dose response and protection against heterologous challenge. Vaccine 15:1834–1838.

van Cuyck-Gandre H, Zhang HY, Tsarev SA, Clements NJ, Cohen SJ, Caudill JD, Buisson Y, Coursaget P, Warren RL, Longer CF. 1997. Characterization of hepatitis E virus (HEV) from Algeria and Chad by partial genome sequence. J Med Virol 53:340–347.

Yin S, Purcell RH, Emerson SU. 1994. A new Chinese isolate of hepatitis E virus: Comparison with strains recovered from different geographical regions, Virus Genes 9:23–32.

Zharkikh A, Li WH. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. Mol Biol Evol 9:1119–1147.